IPW for missing data: Taylor series expansion with adjustment for unknown weights

Tanayott Thaweethai

October 21, 2019

1 Setting and procedure

We are interested in fitting a linear regression of Y on X. However, some subjects do not have complete data on X and Y and so the analysis model cannot be fit for all subjects. R is a binary indicator of being a complete case. We assume that missingness is explained completely by Z, a vector of fully observed variables. We fit a logistic regression of R on Z, obtaining $\hat{\alpha}$, and then calculate $\hat{\pi} = Pr(R = 1|Z; \hat{\alpha})$ for all subjects. We then fit the analysis model among subjects with $R = 1$, where each subject is weighted by $1/\hat{\pi}$. We obtain $\hat{\beta}$, the parameter of interest. This is called inverse probability weighting.

2 Notation

Variable	**Dimension**	**Description**
Y_i	Scalar	Outcome for i^{th} subject
X_i	$1 \times p$	Analysis model covariates for i^{th} subject
R_i	Scalar	Binary indicator of whether i^{th} subject is a complete case
Z_i	$1 \times q$	Selection model covariates for i^{th} subject
β	$p \times 1$	Analysis model regression coefficient
α	$q \times 1$	Selection model regression coefficient

3 Inference

3.1 Estimating equations

To estimate α, fitting a logistic regression is equivalent to solving the estimating equations

$$\sum_{i=1}^{N} S_{\alpha,i}(\alpha) = \sum_{i=1}^{N} Z_i^\top \left(R_i - \frac{\exp(Z_i\alpha)}{1 + \exp(Z_i\alpha)} \right) = 0$$

to obtain $\hat{\alpha}$. We assume $\hat{\alpha}$ converges to α^* as $N \to \infty$. Then, to estimate β, fitting a linear regression with inverse probability weights as a function of $\hat{\alpha}$ is equivalent to solving the estimating equations

$$\sum_{i=1}^{N} S_{\beta,i}(\hat{\alpha}, \beta) = \sum_{i=1}^{N} \frac{R_i}{\pi(\hat{\alpha})} X_i^\top (Y_i - X_i\beta) = 0$$

to obtain $\hat{\beta}$. Again we assume $\hat{\beta}$ converges to β^* as $N \to \infty$.

3.2 Taylor series expansion

Expanding $N^{-1}\sum_{i=1}^{N} S_{\alpha,i}(\hat{\alpha})$ about α^*, and multiplying by $N^{1/2}$, we obtain

$$0 = N^{-1/2}\sum_{i=1}^{N} S_{\alpha,i}(\alpha^*) + N^{1/2} E\left[\frac{\partial S_{\alpha,i}(\alpha)}{\partial \alpha^\top}\right]_{\alpha=\alpha^*}(\hat{\alpha}-\alpha^*) \tag{1}$$

Then expanding $N^{-1}\sum_{i=1}^{N} S_{\beta,i}(\hat{\alpha},\hat{\beta})$ about (α^*,β^*), we obtain

$$0 = N^{-1/2}\sum_{i=1}^{N} S_{\beta,i}(\alpha^*,\beta^*) + N^{1/2} E\left[\frac{\partial S_{\beta,i}(\alpha,\beta^*)}{\partial \alpha^\top}\right]_{\alpha=\alpha^*}(\hat{\alpha}-\alpha^*) + N^{1/2} E\left[\frac{\partial S_{\beta,i}(\alpha^*,\beta)}{\partial \beta^\top}\right]_{\beta=\beta^*}(\hat{\beta}-\beta^*) \tag{2}$$

We define

$$\tau = -E\left[\frac{\partial S_{\beta,i}(\alpha^*,\beta)}{\partial \beta^\top}\right]_{\beta=\beta^*}, \quad \delta = -E\left[\frac{\partial S_{\beta,i}(\alpha,\beta^*)}{\partial \alpha^\top}\right]_{\alpha=\alpha^*}, \quad I_\alpha = -E\left[\frac{\partial S_{\alpha,i}(\alpha)}{\partial \alpha^\top}\right]_{\alpha=\alpha^*}$$

Substituting in these quantities and rearranging the terms in equations 1, we obtain

$$0 = N^{-1/2}\sum_{i=1}^{N} S_{\alpha,i}(\alpha^*) - N^{1/2} I_\alpha(\hat{\alpha}-\alpha^*) \tag{3}$$

$$N^{1/2}(\hat{\alpha}-\alpha^*) = N^{-1/2}\sum_{i=1}^{N} I_\alpha^{-1} S_{\alpha,i}(\alpha^*) \tag{4}$$

Now substituting in the above quantities and rearranging the terms in equation 2, we obtain

$$0 = N^{-1/2}\sum_{i=1}^{N} S_{\beta,i}(\alpha^*,\beta^*) - N^{1/2}\delta(\hat{\alpha}-\alpha^*) - N^{1/2}\tau(\hat{\beta}-\beta^*) \tag{5}$$

$$N^{1/2}(\hat{\beta}-\beta^*) = \tau^{-1}\left[N^{-1/2}\sum_{i=1}^{N} S_{\beta,i}(\alpha^*,\beta^*) - N^{1/2}\delta(\hat{\alpha}-\alpha^*)\right] \tag{6}$$

Then substituting in from equation 4,

$$N^{1/2}(\hat{\beta}-\beta^*) = \tau^{-1}\left[N^{-1/2}\sum_{i=1}^{N} S_{\beta,i}(\alpha^*,\beta^*) - \delta N^{-1/2}\sum_{i=1}^{N} I_\alpha^{-1} S_{\alpha,i}(\alpha^*)\right] \tag{7}$$

$$= \tau^{-1} N^{-1/2}\sum_{i=1}^{N}\left[S_{\beta,i}(\alpha^*,\beta^*) - \delta I_\alpha^{-1} S_{\alpha,i}(\alpha^*)\right] \tag{8}$$

3.3 Central limit theorem

By the central limit theorem,

$$N^{1/2}(\hat{\beta}-\beta^*) \to \text{Normal}(0,\Omega) \tag{9}$$

where

$$\Omega = \tau^{-1}\text{Var}\left[S_{\beta,i}(\alpha^*,\beta^*) - \delta I_\alpha^{-1} S_{\alpha,i}(\alpha^*)\right]\tau^{-\top} \tag{10}$$

3.4 Expression for plug-in estimator

By Slutsky's theorem, we can estimate Ω with $\hat{\Omega}$ where

$$\hat{\Omega} = \hat{\tau}^{-1}\left\{N^{-1}\sum_{i=1}^{N}\left[S_{\beta,i}(\hat{\alpha},\hat{\beta}) - \hat{\delta}\hat{I}_\alpha^{-1} S_{\alpha,i}(\hat{\alpha})\right]^{\otimes 2}\right\}\hat{\tau}^{-1}$$

$$\hat{\tau} = -\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial S_{\beta,i}(\hat{\alpha},\beta)}{\partial \beta^\top}\right]_{\beta=\hat{\beta}}, \quad \hat{\delta} = -\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial S_{\beta,i}(\hat{\alpha},\beta)}{\partial \alpha^\top}\right]_{\alpha=\hat{\alpha}}, \quad \hat{I}_\alpha = -\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial S_{\alpha,i}(\alpha)}{\partial \alpha^\top}\right]_{\alpha=\hat{\alpha}}$$

where $A^{\otimes 2} = AA^\top$. Then, for a given estimator $\hat{\beta}$. we estimate the variance of $\hat{\beta}$ by $\widehat{\text{Var}}(\hat{\beta}) = \hat{\Omega}/N$.

3.5 Naive robust sandwich estimator

If we were to ignore the fact that the weights were estimated (i.e., that α is known), then if we were to expand $N^{-1}\sum_{i=1}^{N} S_{\beta,i}(\alpha,\hat{\beta})$ about β^*, we obtain

$$\begin{aligned} 0 &= N^{-1/2}\sum_{i=1}^{N} S_{\beta,i}(\alpha,\beta^*) + N^{1/2} E\left[\frac{\partial S_{\beta,i}(\alpha,\beta)}{\partial \beta^\top}\right]_{\beta=\beta^*}(\hat{\beta}-\beta^*) && (11) \\ &= N^{-1/2}\sum_{i=1}^{N} S_{\beta,i}(\alpha,\beta^*) - N^{1/2}\tau(\hat{\beta}-\beta^*) && (12) \\ N^{1/2}(\hat{\beta}-\beta^*) &= \tau^{-1}N^{-1/2}\sum_{i=1}^{N} S_{\beta,i}(\alpha,\beta^*) && (13) \end{aligned}$$

Then by the central limit theorem,

$$N^{1/2}(\hat{\beta}-\beta^*) \to \text{Normal}(0,\tilde{\Omega}) \tag{14}$$

where

$$\tilde{\Omega} = \tau^{-1}\text{Var}\left[S_{\beta,i}(\alpha,\beta^*)\right]\tau^{-\top} \tag{15}$$

and we estimate $\tilde{\Omega}$ by $\hat{\tilde{\Omega}}$ where

$$\hat{\tilde{\Omega}} = \hat{\tau}^{-1}\left\{N^{-1}\sum_{i=1}^{N}\left[S_{\beta,i}(\alpha,\hat{\beta})\right]^{\otimes 2}\right\}\hat{\tau}^{-1}$$

and $\hat{\tau}$ was defined as above. This is known generally as the **robust sandwich estimator**. This is the same as the estimator provided in Section 3.4 except for the use of α instead of $\hat{\alpha}$ and subtracting off the term $\hat{\delta}\hat{I}_\alpha^{-1}S_{\alpha,i}(\hat{\alpha})$. By the form of $\hat{\delta}$ and $\hat{I}_\alpha$, we know that $\hat{\delta}\hat{I}_\alpha^{-1}S_{\alpha,i}(\hat{\alpha})$ is nonnegative. Therefore, $\hat{\tilde{\Omega}} \geq \hat{\Omega}$, and so accounting for estimation of the weights results in a lower standard error estimate. In other words, assuming the weights are known results in an overestimate of the true standard error if the weights were in fact estimated.

3.6 Plug-in estimator for logistic selection model and linear analysis model

$$S_{\alpha,i}(\hat{\alpha}) = Z_i^\top\left(R_i - \frac{\exp(Z_i\hat{\alpha})}{1+\exp(Z_i\hat{\alpha})}\right), \quad S_{\beta,i}(\hat{\alpha},\hat{\beta}) = \frac{R_i(1+\exp(Z_i\hat{\alpha}))}{\exp(Z_i\hat{\alpha})}X_i^\top\left(Y_i - X_i\hat{\beta}\right) \tag{16}$$

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i(1+\exp(Z_i\hat{\alpha}))}{\exp(Z_i\hat{\alpha})}X_i^\top X_i, \quad \hat{\delta} = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{\exp(Z_i\alpha)}X_i^\top\left(Y_i - X_i\hat{\beta}\right)Z_i, \tag{17}$$

$$\hat{I}_\alpha = \frac{1}{N}\sum_{i=1}^{N}\frac{\exp(Z_i\hat{\alpha})}{\left[1+\exp(Z_i\hat{\alpha})\right]^2}Z_i^\top Z_i \tag{18}$$